United States Department of Agriculture

Service Center Initiative (SCI)

Standard for Service Center Data Naming

Prepared by
Service Center Data Team

Abstract: Standard for creating names for data objects to be stored in the central metadata repository. Rules are outlined for naming business data objects (those described through logical data models and associated dictionaries) and technical data objects associated with physical data models, data files and databases. The naming standards described in this document apply to all new development and redesign initiatives.

Keywords: standard, data, naming, metadata

Published and distributed in August 6, 1999

(Revision of Draft Service Center Data Naming Standard, August 1998)

Introduction

(This introduction is not part of SCI Std 009-01, Standard for Service Center Data Naming.)

Standards and guidelines for the development of business and technical names will help promote a common understanding of data within USDA. By developing standardized and descriptive names, system developers and business users will be able to locate, reuse, merge and analyze data from other systems to meet their information needs. Reusing data can reduce redundant data stores and data collection applications, and can improve the quality and utility of newly developed applications.

Developing and implementing naming standards is also a prerequisite to the successful implementation of the Service Center's central metadata repository.

The Service Center Data Team (Data Team) developed the material contained herein. This team is composed of representatives from Natural Resources Conservation Service (NRCS), Farm Service Agency (FSA), and Rural Development (RD). Sponsorship and direction for the Data Team comes from the Executive Director of the National Food and Agriculture Council (NFAC) of the United States. Department of Agriculture (USDA) and the partner Service Center agency's Chief Information Officers (CIOs).

The individuals who contributed to the development of this guide are:

Jim Schmidt, Leader

Tom Richards	Richard Verhoff	Dave Butler
Chuck Calcaterra	Lee Stieger	Dorn Egley
Jerry Kleiman	Mindy Gehrt	Norma Westbrook

Figure 1—Working group list

Table of contents

STANDARD FOR SERVICE CENTER DATA NAMING

1. Overview

All data objects for new and reengineered information technology (IT) systems shall be captured in a central metadata repository. This repository shall support centralized data management and administration, with the overall goal of improving communication and sharing of enterprise business information across business areas and IT systems. In order to realize this vision, the metadata in this central repository must be recognizable, understandable and comparable across the various sources. Developing and applying common names for data objects to be stored in the central metadata repository facilitates common understanding of data.

1.1. Scope

The data object naming standards described in this document apply to metadata created through all new IT development initiatives, as well as redesign/reengineering projects.

NOTE–There are no current requirements for standard names to be developed for legacy data (created and maintained through pre-existing application systems); agencies interested in documenting legacy data in the central metadata repository have the option of creating standard names or using the existing legacy names.

IT applies both business (logical) names and technical (physical) names for data objects associated with data models, data dictionaries, data files, databases, etc.

1.2. Purpose

The purpose of this naming standard is to provide developers, data analysts and stewards with the rules for developing common names. The use of common names will improve understanding of data, promote data quality, reduce the occurrence of aliases and minimize data redundancy.

1.3. Definitions

The following definitions apply to key terms used throughout the remainder of this document.

1.3.1. Business name

An English-like name that describes a single data object (i.e. data model, database, file, table, or data element). Also, referred to as the "conceptual" or "logical" name, business names must be meaningful to the business community, and must reflect the definition or true meaning of the data object.

1.3.2. Category identifier

An identifier used to classify data elements into specific categories. Each business name must include a category identifier. It is normally the last term in the business name, following the prime word and its modifiers. However, the category identifier may be omitted if its usage would cause awkward redundancy in the business name. For example, *Forestland Acreage* versus *Forestland Acreage Quantity,* or *Crop Yield* versus *Crop Yield Quantity.* See Appendix B for a list of valid category identifiers.

1.3.3. Modifier

A term that further refines the prime word. Modifiers may be used with category identifier and prime words to more accurately describe the data object.

1.3.4. Prime word

Identifies the subject to which the business name refers. It may consist of a single word, compound word, or, in rare cases, an acronym.

1.3.5. Technical name

The actual name of the data object as it is represented in a physical database, file or application program. Technical names are created based on the business name, but are constrained by the requirements of the specific database management system (DBMS), programming language or other technology tool used to support the data. Such constraints may include the length of the name, character that may be used to separate terms (such as spaces or underscores), whether upper or lowercase characters should be used, reserved words or characters that may not be used, etc.

1.4. Acronyms and abbreviations

CASE	computer-aided software engineering
DBMS	database management system
IT	information technology
URL	universal resource locator

2. Standard name requirements and conventions

The metadata required for each data object includes its

– Definition
– Business name
– Technical name, if applicable

NOTE–Several other pieces of information (such as datatype, domain, etc. may also be required in the minimum metadata set), but these are outside of the scope of this naming standard.

2.1. Business Name

The Business name is an English-like name that describes a single data object (i.e. data model, database, file, table, or data element). Also referred to as the "conceptual" or "logical" name, business names must be meaningful to the business community, and must reflect the definition or true meaning of the data object.

Business names are developed by business users or data stewards, often with the assistance of data analysts and/or other members of development teams.

The terms or words that comprise a business name should be fully spelled out, with individual words separated by spaces. Business names must consist of a prime word, a category identifier, and optionally, one or more modifiers. Although there may be several technical names for a data object (for example, the data object may be included in multiple databases and be named differently in each), there should be only one business name representing each individual concept. In other words, if the definition is the same, the business name should also be the same. Cross-referencing of technical names to business names shall be managed through the central metadata repository.

A single business name should be constructed based on the business definition of the data item. The following semantic, structure and syntax conventions shall be used in developing a business name.

2.1.1. Business name semantic rules

The following rules apply to the creation of business names:

1. Business names must be derived or based on their definition. The business name is directly linked to its definition. Two data objects with different definitions cannot have the same name.
2. Business names must be clear, accurate and self-explanatory.
3. Business names must not contain the names of organizations, computer or information systems, directives, forms, rows or columns of screens or reports.

4. Business names must not express multiple concepts, either implicitly or explicitly. Do not use overly generic names, such as "notes", "value", “screen display”, etc. as a business name for a specific data object.

2.1.2. Business name structure rules

Business names should be constructed based on the following structure rules:

1. Business names consist of a required prime word, optional modifiers and normally a category identifier (for data elements) structured as follows:

Table 1–Business name structure

Modifier	Prime word	Modifier(s)	Category identifier	Modifier
Optional	required	Optional	required (for data elements)	Optional

2. The prime word should unambiguously identify the object to which the business name refers.
3. Modifiers further refine the prime word. For example, an organization may be interested in information about two distinct codes for crop insurance – Crop Insurance Availability Code and Crop Insurance Carrier Code. The modifiers "availability" and "carrier" are used to distinguish between the two types of Crop Insurance Codes.
4. Modifiers should be used only as necessary.
5. A category (sometimes referred to as class) identifier is used to designate the general category of data described by a business name. See Appendix A for a list of authorized category identifiers.
6. Category identifiers are not mandatory when their usage would cause awkward redundancy.
7. Category identifiers are normally the last word of a business name.
8. Category identifiers are reserved words and are not to be used as prime words.

2.1.3. Business name syntax rules

Business names must be formatted based on the following syntax rules:

1. Business names shall be constructed of upper and lower case letters, using initial capitalization, such as in a title. (e.g., Phone number).
2. Business names must not exceed 50 characters in length.
3. Abbreviations and acronyms should be avoided. Exceptions to this rule include universally accepted abbreviations or acronyms (e.g., FAX for Facsimile). A list of standard acronyms is available at the the following universal resource locator (URL) address: http://www.fsa.usda.gov/dam/kcmo/dditst/imsdacro.htm .

4. Only alphabetic characters (A-Z, a-z), spaces, in rare cases numbers (0-9) and hyphens are permitted.
5. Each word of a business name is separated by a space or by using a hyphen to connect multiple words (e.g., Loan Write-off Amount).
6. Prepositions (e.g., at, by, for, from, in, of, to) are not permitted in a business name unless required for clarity (e.g., Power of Attorney Code).
7. Articles (e.g., a, an, the, etc.) are not permitted.
8. Conjunctions (e.g., and, or, but) are not permitted.
9. Verbs should be avoided.
10. When naming a many-to-many relationship, the names of the two entities should generally be combined to form the name. For example, the link between customer and office would be called *Customer Office*.
11. When naming a recursive relationship, the word "Association" should be added to the name of the entity. For example, a recursive relationship for *Customer* (i.e., a customer to customer relationship) would be named *Customer Association*.

2.2. Technical name

The technical name is the actual name of the data object as it is represented in a physical database, file or application program. Technical names are created based on the business name, but are constrained by the requirements of the specific database management system (DBMS), programming language or other tool used to support the data. Such constraints may include the length of the name, character that may be used to separate terms (such as spaces or underscores), whether upper or lowercase characters should be used, reserved words or characters that may not be used, etc.

Since the same data objects may be used across many IT systems, they may be represented by multiple technical names. However, if the same concept is represented (i.e., the meanings behind those data objects is identical) they should map back to the same business name.

Standard technical names are created based on the standard business name. They are constrained by specific requirements of the technology tools being implemented, and thus must comply with those requirements as well. The following conventions apply to the creation of standard technical names:

1. Technical names are constructed using the business name and abbreviating as necessary to meet length restrictions. If length restrictions can be met without abbreviating, the technical name may use the full spelling.
2. Multiple technical names (logical, physical, etc.) of varying lengths may exist to conform to the Computer-aided Software Engineering (CASE) tool, language, or DBMS being used. All technical names shall be linked to a single business name and a single definition in the metadata repository.

3. If the business name must be shortened to meet physical length restrictions, standard abbreviations must be used if available. For a complete listing of approved standard abbreviations refer to the following URL address: http://wwwaix.fsa.usda.gov/imsdabr.htm.
4. For words not on the standard list, follow the abbreviation procedure presented in Appendix B.
5. When needed, words in technical names shall be abbreviated starting with the last word in the name until the length restriction is met. For data elements, the category identifier is usually the first word to be abbreviated. See Appendix A for a list of standard category identifier abbreviations.
6. When a technical name is designed from standard abbreviations and the length still exceeds the technical constraint, drop the least meaningful component of the name until the length restriction is met. One option is to eliminate the Category Identifier first and then the next least meaningful word while trying to create an abbreviation for the business name that is recognizable.
7. Delimiters between each word may be used as dictated by the programming language or technology tool. The tool or language may also determine upper/lower-case character constraints.

Appendix A–Category identifier definitions

Table A.1—Category identifier definitions

Category	Identifier	Description and/or definition structure
Abbreviation	ABR	A shortened form of a written word that is used in place of the whole word.
Amount	AMT	A monetary value. (Includes: average, balance deviation, factor, index, level, mean, mode, scale, and yield)
Angle	ANG	The rotation measurement between two lines and/or planes diverging from a common point and/or line. (Includes azimuth and heading)
Area	AR	The measurement of a surface expressed in unit squares (2 dimensional).
Code	CD	A combination of one or more numbers, letters, or special characters substituted for a specific meaning. Represents finite predetermined values. (Must have a specific domain.) (Includes: Category and status)
Coordinate	CN	Designation of the location of a line or plane. (Includes: latitude and longitude)
Count	CT	The total number of individual entities involved in a given unit or sample.
Date	DT	The designation of a specific 24 hour period of time.
Dimension	DIM	A measured linear distance (1-dimensional). (Includes: altitude, depth, diameter, elevation, height, length, radius, vertex, and width.)
Identifier	ID	A combination of one or more numbers, letters, or special characters which designate a specific object/entity, but which have no readily definable meaning. (Must have a general domain) (Includes: designator, key, number
Indicator	IND	Similar to code and contains only two possible conditions (e.g., on/off, yes/no, etc). An indicator is always 1 character.
Mass	MS	The measure of inertia of a body.
Name	NM	A designation of an object and/or entity expressed in a word or phrase.
Number	NBR	A synonym for "Identifier" which may be used when common usage is "Number" rather than "Identifier" (e.g., Social Security Number, Phone Number, etc).
Percent	PCT	Per hundred, in, to, or for every hundred.
Quantity	QTY	A non-monetary numeric value. (Includes: average balance, count, deviation, factor, index, level, mean, median, mode, and scale).
Rate	RT	A quantity or degree of something in relation to units of something else (for example: bushels per acre). (Includes: acceleration, density, factor, flow, force, frequency,

Category	Identifier	Description and/or definition structure
		humidity, impedance, inductance, intensity, magnitude, moment, percent, power, pressure, resistance, scale, speed, tension, torque, velocity, viscosity, and voltage.)
Text	TXT	An unformatted character string, generally in the form of words. (Includes: category and comments.)
Time	TM	A designation of a specified chronological point within a period.
Volume	VOL	Measurement of space occupied by a three-dimensional figure as measured in cubic units.
Weight	WT	The force with which an object is attracted toward the earth and/or another celestial body by gravitation.

Appendix B–Guidelines for abbreviating names

Business names are generally abbreviated to derive the technical name. They should be abbreviated from right to left (i.e., abbreviate the word at the end of the name first, then the next word to the left and so on until the name meets the length restriction). Delimiters between each word may be used as dictated by the programming language or CASE tool (e.g., use hyphens or underscores as dictated by the programming language). Delimiters for the Business Name shall always be a space. Standard abbreviations must be used when available. If a standard abbreviation for a word is not found in the listing at the URL address http://wwwaix.fsa.usda.gov/imsdabr.htm, use the following steps to generate a new abbreviation. New abbreviations should be submitted to a data representative for inclusion in the standard list.

B.1. Abbreviation rules

The basic (i.e., root) form of each word shall be abbreviated and used for all extensions of the basic word. The basic form of the word Builder and Building is Build, therefore the abbreviation for Build (BLD) shall be used for all three words.

Determine if the basic form of the word is long enough to be abbreviated, i.e. greater than four characters. Do not abbreviate words with four or fewer characters. Exceptions are noted in the Standard Abbreviation Listing, which is available at the URL address http://wwwaix.fsa.usda.gov/imsdabr.htm.

– Apply the following abbreviation guidelines, using the basic form of each word:

– Abbreviated words must be in the same sequence as the words in the business name.

– The abbreviations usually begin with the same letter as the word being abbreviated. The order of letters in the abbreviation shall parallel the sequence of letters in the word.

– Generally, eliminate the vowels from a word to form the abbreviation. If the word begins with a vowel, the initial vowel should be retained.

– Generally, if a double consonant appears in the abbreviation, drop one of the consonants (e.g., Commodity => CMMDTY => CMDTY, Pattern => PTTRN => PTRN). Exceptions are allowed for common abbreviations used across USDA (e.g., COLL for Collect and COOP for Cooperative).

– If the abbreviation contains a "ck" drop the "c" (e.g., Brick => BRCK => BRK).

– Abbreviations can, in some cases, duplicate existing abbreviations. It is not necessary to establish a unique abbreviation for every word. If two words share the same abbreviation but are not likely to be used in the same data item, they may share an identical abbreviation (e.g., FERT may be used for Fertilizer and Fertility, but are unlikely to be used together in the same data item). The abbreviation used in the

context of the full business name, or in the context of the database table or file where it is found, should resolve any ambiguity.

– Do not abbreviate acronyms regardless of their length.

– Abbreviations of a single word shall not contain hyphens or other special characters. If the word normally contains a hyphen, drop the hyphen and derive an abbreviation for the concatenated word. For example, to abbreviate the word "in-transit", drop the hyphen and develop an abbreviation for "intransit".

– Avoid lengthy abbreviations for words. Abbreviation lengths of 4 or fewer characters are ideal whenever practical. It is important, though, to have abbreviations that are meaningful, promote clarity and are more than an indecipherable jumble of letters.

– Often, technical names are constrained by the tools used (such as the 18-character limit imposed by some DBMS's). These constraints require close attention to the length of individual word abbreviations. The average data element business name of 3 to 4 words can quickly exceed the character length limit when abbreviations for individual words exceed 3 to 4 characters.

– Any words that have well known and widely understood initials should be referenced using those initials (e.g., Accounts Receivable is AR).

– Repeatability and consistency of abbreviation must be promoted across systems. New abbreviations should be added to the standard abbreviation list.

B.2. Exceptions to the abbreviation rules

Generally, exceptions can be granted under the following circumstances:

– When common usage of an abbreviation exists (e.g., lb for pound).

– For clarity (e.g., Maximum is more commonly recognized as MAX rather than MXM).

– To avoid awkwardness, as illustrated in the following example:

> The Business Name 'Organization Classification Level Name' can be abbreviated by dropping vowels as: ORGNZTN_CLSFCTN_LVL_NM.
>
> While it follows the general rules for abbreviating, this example is awkward and may exceed the length limit. A better physical name abbreviation that still conveys the correct meaning would be 'ORG_CLS_LVL_NM'.

– When the abbreviation may be confused with other abbreviations or names. For example, Color and Clear could both be abbreviated CLR. An exception to the normal

abbreviation rules may be made for Color, since CLR is more commonly accepted for Clear.

– When the possibility exists that two or more data objects will end up with the same name in a situation where uniqueness is essential. For example, Employee and Employer would both use the abbreviation of the base word Employ (EMP). The two data elements Employee Name and Employer Name would have the same technical name abbreviation (EMP_NM), and could reside in the same table causing a collision of names.

– Avoid abbreviating obscure names, if possible (e.g., Coppice or Humus).

– When the business name is a chemical formula (e.g., Calcium Carbonate - $CACO_3$, Calcium Nitrate - $CANO_3$).

– When the abbreviation does not adequately represent the full word (e.g., Hedging would be abbreviated as HDG, however HEDG would more adequately represent the full spelling of the business name).